AD_____

Award Number: DAMD17-97-1-7122

TITLE: Cell Type-Specific mRNA Amplification and Expression Profiling from Breast Tumor Sections

PRINCIPAL INVESTIGATOR: Donna G. Albertson, Ph.D.

CONTRACTING ORGANIZATION: University of California-San Francisco
San Francisco, California 94143-0962

REPORT DATE: December 2001

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 2001	Final (20 Sep 97 - 15 Nov 01)

4. TITLE AND SUBTITLE
Cell Type-Specific mRNA Amplification and Expression Profiling from Breast Tumor Sections

5. FUNDING NUMBERS
DAMD17-97-1-7122

6. AUTHOR(S)
Donna G. Albertson, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California-San Francisco
San Francisco, California 94143-0962

E-Mail: albertson@cc.ucsf.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Report contains color

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The evolution of solid tumors involves acquisition of genetic abnormalities, which result in changes in both the set of genes expressed and the relative levels of gene expression. However, the increasing number of candidate genes whose expression needs to be evaluated for prognostic, diagnostic, therapeutic, or research purposes will require obtaining material from numerous tissue sections. Therefore this proposal is motivated by the need for more effective use of clinical specimens, and will address the problem of obtaining sufficient and cell type specific mRNA from clinical breast tumor specimens. This will entail adapting/developing procedures to amplify with fidelity the mRNA repertoire expressed in small numbers of normal, pre-cancerous and malignant breast epithelia. To this end, in this project period, we have concentrated effort on establishing and validating microarray-based assays for measuring gene expression levels and have demonstrated the capability to isolate and amplify mRNA from cultured cells. Realization of these objectives will allow, in the future, development of a resource, consisting of amplified mRNA populations from individual cells from normal and tumor material, that can be used for evaluation of the prognostic, diagnostic and/or therapeutic importance of genes expressed in breast cancer.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	14
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION:

The evolution of solid tumors involves acquisition of genetic abnormalities, which result in changes in both the set of genes expressed and the relative levels of gene expression. Therefore it is desirable to be able to characterize and compare the levels of expression of genes in normal and tumor tissue. Currently, DNA microarray technology allows expression of tens of thousands of genes to be assayed simultaneously. However, if tissue samples are small, it may be difficult to extract sufficient mRNA from the specific cell types under study to perform these measurements. Therefore in order to obtain expression profiles from small specimens, methods are being evaluated that will allow the efficient and democratic amplification of the mRNA repertoire of the cells. Since microarrays provide an efficient methodology for assessment of gene expression, this project has focused on evaluation and optimization of methods for both labeling mRNA for microarray hybridization and for amplification of the starting material. This work has been carried out in parallel with the development of highly quantitative microarray-based methods for the analysis of DNA copy number by comparative genomic hybridization (array CGH). The array CGH studies have facilitated the expression array work in this project, since, in contrast to expression measurements, in array CGH the relative copy numbers of different nucleic acids species often are known, which allows accurate assessment of measurement performance. Validation of the different procedures for amplification of RNA and measurement of expression levels has relied primarily on measurement of relative abundance of transcript levels of test and reference populations using microarrays capable of measurement sensitivity over more than two orders of magnitude. When comparing two RNA populations (e.g. one having been amplified and one representing the starting material), uniformity of labeling or equal representation of genes in the two mRNA populations is indicated by constant intensity ratios across all array spots over the entire abundance range. Deviant ratios indicate non-uniform amplification of those particular genes. Quantitative PCR using the Taqman system has also been used as a validation procedure on selected genes. Development of procedures that provide efficient and democratic amplification of the mRNA repertoire of cells will allow more efficient use to be made of valuable clinical samples.

BODY:

A. Reporting Period.

This final report summarizes work undertaken during the entire award period. Work focused primarily on Technical Objective1, Tasks1-3 and 7 with the goal of developing and validating labeling protocols and assay procedures.

Statement of Work

1. Technical Objective 1. Demonstrate linear amplification of high complexity aRNA (amplified antisense RNA) from a homogeneous population of cells, the breast tumor cell line BT474.

Task 1	Months 1-3	Validate assay for expression levels. Make test RNA population by transcribing test genes *in vitro*, label and hybridize to array of test clones. Demonstrate linearity of the assay.
Task 2	Months 4-9	Grow BT474 cell cultures, isolate mRNA, measure expression levels of test genes in mRNA isolated from BT474 and estimate complexity by hybridization to IMAGE cDNA array.

Task 3	Month 4-9	Carry out amplification on various amounts of bulk BT474 mRNA down to 0.1 pg, measure expression levels of test genes in aRNA and hybridize to IMAGE cDNA array.
Task 4	Months 10-13	Prepare frozen and fixed samples of BT474 cells and cut sections. Carry out mRNA amplification on sections of BT474 cells, measure expression levels of test genes in aRNA and estimate complexity by hybridization to IMAGE cDNA array.
Task 5	Months 10-13	Prepare fluorescently labeled probes for test genes and carry out *in situ* hybridization to sections of BT474 cells. Measure intensity of fluorescent hybridization signals and determine relative levels of expression of test genes in the cells in the sections.
Task 6	Months 10-13	Compare relative levels of expression of test genes in aRNA and *in situ* as determined by FISH in Task 7. Compare results of hybridization to IMAGE cDNA array with aRNA made to whole sections and microdissected cells.
Task 7	Months 10-13	Optimize protocols (Tasks 4-6).

2. Technical Objective 2. Apply the techniques from Objective 1 to amplify mRNA from microdissected cells from frozen and formalin fixed sections containing normal ductal epithelial cells, DCIS and invasive carcinoma. Use this material to obtain expression profiles for these different cell types using SAGE and hybridization to an array of clones from the IMAGE cDNA library.

Task 8	Months 14-24	Prepare fluorescently labeled probes for test genes and carry out *in situ* hybridization to breast tumor sections. Measure intensity of fluorescent hybridization signals and determine relative levels of expression of test genes in different cell types in the tumor section.
Task 9	Months 14-24	Carry out mRNA amplification on tumor sections.
Task 10	Months 14-24	Measure expression levels of test genes in aRNA from tumors.
Task 11	Months 14-24	Compare relative levels of expression of test genes in aRNA and *in situ* as determined by FISH in Task 8.
Task 12	Months 14-24	Carry out expression analysis on aRNA from tissue sections by hybridizing the aRNA to an array from the IMAGE cDNA library.

Task 1. Validate assay for expression levels.

We used cDNA microarrays to measure and evaluate the linearity and uniformity of the RNA amplification procedure. In this assay, differentially fluorescently labeled probes are made from two nucleic acid populations. One mRNA is labeled with a fluorochrome such as Cy3 and the other with another fluorochrome (e.g. Cy5). These probes are then hybridized to the microarray and the fluorescence intensity of each probe is determined on each array spot. If all genes in the array are represented at equal levels in the two RNA populations, then the fluorescence intensity ratios should be constant across all spots. Deviant ratios indicate non-uniform amplification of those particular genes. We demonstrated that our microarray capability (Pinkel *et al.*, 1998; Snijders et al., 2001; Jain et al., 2001) provided measurement sensitivity

over more than two orders of magnitude and therefore provided the appropriate measurement precision for the analysis of the amplification procedure.

Task 2. Isolate RNA from BT474 cells.

We evaluated the quality and labeling of the RNA prepared using several commercially available kits for total RNA isolation and/or mRNA isolation. The RNA isolation appeared to affect the amount of background fluorescent hybridization signal on the microarray substrate. We found that extraction of total RNA using Trizol (BRL) followed by isolation of mRNA using either the Promega or Invitrogen systems yielded good signal and low background. We also modified the Trizol RNA isolation procedure to obtain DNA from the same samples that was suitable for use for array CGH, thereby providing the capability to make both expression and DNA copy number measurements on the same samples (see Appendix).

Task 3. Carry out amplification on various amounts of bulk BT474 mRNA.

The aRNA amplification procedure (Eberwine et al., 1992) was carried out using RNA from MCF7 cells. Approximately 100-fold amplification of the mRNA was achieved with a starting concentration of 200 ng, while amplification was less efficient with greater amounts of starting RNA. Further amplification was possible by performing a second round of synthesis, but the cDNA was significantly decreased in size, indicating that the second round of synthesis failed to preserve the representation of the various RNA species in the population.

Task 7. Optimize protocols.

Linear amplification methods allow detection of gene expression from smaller amounts of starting material. However, they must be assessed for biases introduced during the process. We investigated several sources of bias, including those introduced during the amplification procedure itself, during the labeling reaction and in the array measurements.

a. Comparison of two different labeling methods. We evaluated direct labeling in which Cy3 or Cy5 labeled nucleotides are incorporated into the probe during the reverse transcriptase (RT) reaction and the use of amino-allyl labeling, an indirect method. For amino-allyl labeling the amino-allyl modified nucleotide is incorporated into both the test and reference by the RT reaction and subsequently the two populations are differentially labeled with Cy3 and Cy5 by a chemical reaction. The potential advantage of the indirect method is that the RT enzyme is incorporating the same nucleotide into the test and reference populations, and therefore may avoid biases due to different efficiencies of incorporation into specific sequences. The amino allyl labeling method was demonstrated to provide more uniform labeling of the samples compared to incorporation of fluorescently labeled nucleotides (see Appendix). This protocol is currently being used for hybridization to arrays comprised of 70-mer oligonucleotides (Operon). It is also being used with arrays comprised of small subclones of genomic DNA from regions recurrently amplified in breast cancer. These arrays are used to identify all expressed sequences in the region.
b. Reproducibility of amplification procedure. Some inter-experiment variation in the amplification procedure was found by comparing independent amplifications of an RNA population. Greater variations in the ratios were observed in comparison to variation between self vs. self comparisons.

c. Degree of preservation of relative gene expression in amplified vs. unamplified RNA populations. The affect of amplification on the measured relative levels of expression of ERBB2 in two cell lines was determined by measuring expression levels in amplified and un-amplified RNA. The levels were similar, indicating that amplification had allowed qualitatively preserved relative gene expression.
d. Comparison of array-based methods with quantitative RT-PCR (TaqMan). Probes for ~20 genes were prepared, the expression of these genes was determined by quantitative RT-PCR and compared to relative levels determined in the array experiments. These comparisons found a considerable discrepancy in the measurement of abundant genes. For example, the relative levels of expression of ERBB2 in BT474 a high level expresser and MCF7, a low expresser were compared. Array measurements indicated that the level of expression in BT474 is 9 times that in MCF7, while by quantitative RT-PCR ERBB2 expression differed by ~100. Possible causes were investigated, including background signals on array spots and non-specific hybridization due too low stringency in the hybridization. It is likely that non-specific hybridization and improper background subtraction are contributing to suppression of the dynamic range of the array measurement. In parallel experiments studying DNA copy number by array CGH, it was also found that some image acquisition hardware is highly sensitive to variations in slide thickness, which may also introduce significant error in the array measurements.

KEY RESEARCH ACCOMPLISHMENTS:

- Amplification of mRNA
  1. ~100 fold amplification of starting mRNA in a single round of *in vitro* transcription.
  2. Independent amplification reactions of a single RNA population show high correlations, but greater variations in microarray ratios than unamplified self vs. self comparisons.
  3. Demonstration of preservation of relative gene expression levels after amplification. However, expression levels are not quantitatively preserved.

- Microarray-based analysis of gene expression
  1. Demonstration of detection sensitivity in microarray expression measurements over a range greater than two orders of magnitude.
  2. Development of protocols for fluorescent labeling of RNA. Demonstration of 10% standard deviation in self vs. self ratios on microarrays using amino-allyl labeling.
  3. Array-based measurements under-estimate expression levels compared to quantitative RT PCR (TaqMan).
  4. Application of protocols developed using cDNA arrays to expression measurements using arrays comprised of gene-specific oligonucleotides or random subclones of genomic DNA from regions of recurrent amplification in breast cancer.

REPORTABLE OUTCOMES:

Abstracts

- Ylstra, B., Livezey, K. and Albertson, D.G. Identification of Vitamin D 24 Hydroxylase (*CYP24*) as a candidate oncogene by microarray CGH and quantitative expression analysis. American Association for Cancer Research 91st Annual Meeting, April 1-5, 2000.
- Albertson, D., Livezey, K. and Ylstra, B. mRNA amplification and expression profiling using microarrays. Era of Hope Meeting, June, 2000.

Funding applied for using work in this project as preliminary data

- NIH/NCI R33 CA94407-01 (Albertson, P.I.). "Amplicon Profiling by Array CGH." 12/1/01-11/30/04.

Employment

- Kristin Livezey, postdoctoral fellow supported by this project has taken a position with Gen-Probe, San Diego, CA.

CONCLUSIONS:

Work in this project focused on investigating labeling and amplification of RNA by the Eberwine procedure (Eberwine et al, 1992). cDNA microarrays were used to evaluate the linearity and fidelity of the aRNA amplification procedure and for optimization of the aRNA amplification. Amplification of ~100-fold was possible using this approach. However, it was demonstrated that the amplification procedure introduces variation into the measurements. In addition, ratios were not preserved quantitatively. Therefore, for some applications in which qualitative assessment of relative gene expression is sufficient, the procedure will be adequate. However, one should use caution when interpreting microarray data obtained after amplification.

A protocol for labeling using amino allyl modified nucleotides was also developed. It was demonstrated that the amino allyl labeling method introduced fewer biases into the labeling procedure, and therefore it is the preferred method for labeling for microarray analyses.

Array-based expression measurements were compared with measurements made using quantitative RT-PCR (TaqMan). Measurements made by the two methods varied by as much as a factor of 100 in relative expression levels of some genes. The arrays provided an under-estimate of the expression level, which may in part be due to factors such as non-specific binding of label to the array spots and improper estimation of background.

Development of procedures that provide efficient and democratic amplification of the mRNA repertoire of cells will allow more efficient use to be made of valuable clinical samples.

REFERENCES:

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P. 1992. Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. U. S. A. 89, 3010-3114.

Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra,B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel D. and Albertson, D. G. 2001. Assembly of microarrays for genome-wide measurement of DNA copy number by CGH. Nature Genet. 29, 263-264.

APPENDIX:

Protocol 1. CGH quality DNA from Trizol, for frozen breast tumor sections

Bauke Ylstra, Albertson Laboratory UCSF Cancer Center

We estimate 1μg of DNA/ 200 μg of RNA

1. Following phase separation (Trizol protocol procedure step 2),
2. Spin down 12.000g 5min 4 °C.and remove any remaining aqueous phase overlaying the interphase (DNA sits in interphase)
3. Add back extraction buffer (BEB) to interphase organic phase mixture, equal volume.
4. (is equal to 0.5 mL BEB per 1.0 ml of Trizol used for the initial homogenization)
5. Mix vigorously for 15 sec. by inversion
6. Spin 12,000 x g for 15 min 4 °C
7. Transfer upper aqueous phase and save interphase for protein isolation
8. Add 0.8 Volumes of isopropanol (is equal to 0.4 mL isopropanol per 1.0 mL Trizol used for initial homogenization)
9. Mix and incubate 5 min RT
10. Spin 12,000 x g for 15 min 4 °C
11. Wash 70% ETOH dissolve in 1XTE pH 8.0 to resuspend the DNA *(This DNA will perform well in many reactions for CGH we recommend several* Phenol Chloroform Isoamylalcohol *(PCI) steps, easiest using Phase Lock Gels (PLG) by Eppendorf)*
12. Spin down PLG 2ml light or Green tube at 12,000-16,000 g in micro-centrifuge for 20-30 seconds
13. Add 200 μL 1XTE pH 8.0 suspended DNA to the 2 ml PLG light tube. And add equal volume of PCI directly to the tube.
14. Mix the organic and the aqueous phases thoroughly by inverting (Do not Vortex!!!)
15. Centrifuge at 12,000-16,000 g for 5 minutes to separated the phases.
16. Add equal volume of PCI directly to the tube and repeat #12 and #13.
17. Add equal volume of CI directly to the tube and repeat #12 and # 13.
18. Centrifuge at 12,000-16,000 g for 5 minutes to separated the phases.
19. Transfer the aqueous solution to a new 1.5ml micro-centrifuge tube.
20. Add 20 μL 3M sodium acetate pH 7-5.2. Mix and add 2 to 2.5 Volume of 95-100% ETOH. Mix and you can see DNA coming out of the solution.
21. Spin down the DNA in micro-centrifuge at full speed for 10 to 15 minutes.
22. Discard the liquid and add 100 μL icecold 70% ETOH. Vortex the sample and spin down the DNA in micro-centrifuge at full speed for 10 to 15 minutes.
23. Discard the liquid with a pipet making sure no liquid is left in the tube.
24. Resuspend the pellet with TE or water and the DNA is ready to be use.

Back extraction buffer (BEB):

250mL:

4M Guanidine Thiocyanate	FW 118.2,	118.2 g	
50 mM Sodium Citrate NaCi	FW 294.1,	3.68g	(or 12.5 mL of 1M stock)
1 M Tris (free base)Tris	FW 121.14,	30.285g	

Protocol 2. Amino-allyl cDNA Synthesis

Kristin Livezey, Albertson Laboratory, UCSF Cancer Center

RT reaction step:
Adapted Display-THERMO-RT kit (Display Systems Biotech)

DISPLAY THERMO-RT 5X BUFFER		**4.0 UL**
***5MM AA-DNTP MIX**		
2.0 UL		
T25V Primer (10 uM)	2.0 ul	
RANDOM PRIMERS (1UG/UL)		**1.0 UL**
TOTAL RNA		
(10UG)		
Depc H20	to 18.0 ul	
DisplayThermo-RT Initiator mix	1.0 ul	
	18.0 ul	

Incubate at 65 C for 10 min
Decrease temp to 42 C, add 2.0 ul display THERMO-RT Terminator Mix
Incubate 42 C 40 min
65 C 15 min
can store at –20 C

*5mm AA-dNTP Mix	
5 ul	dATP (100 mm) (Gibco)
5 ul	dCTP (100 mm)
5 ul	dGTP (100 mm)
4 ul	AA-dUTP (100 mm)(SIGMA)
1 ul	dTTP (100mm)
80 ul	H20
100 ul RXN	

RNA degradation:

add 2 ul 2.5 M NaOH, incubate 37 C 15 min
to neutralize: add 10 ul 2 M Hepes free acid

probe cleanup:

Fill one Microcon 30 with 470 ul water, add neutralized reaction, spin 12K for 8 minutes.

Dump flow through and refill with water and repeat 1X
Elute
Dry elute in speed vac

Coupling:
RESUSPEND CDNA IN 9 UL 0.05 M NABICARBONATE BUFFER, PH 9.0. LET SIT FOR 10-15 MIN AT ROOM TEMP

Resuspend monofunctional NHS-ester Cy3 or Cy5 dye in 10 ul DMSO,aliquot 1.25 ul X 8 tubes and dry in speedvac(store at 4C)
TRANSFER CDNA + BICARBONATE BUFFER TO ALIQUOT OF DYE

Let incubate 1 hr at RT in dark

Quenching and cleanup
Add 4.5 ul 4 M hydroxylamine, incubate 15 min RT in dark

To remove unicorporated/quenched dyes use Qia-Quick PCR Purification Kit (Qiagen)

Combine cy3 and cy5 rxns
Add 70 ul water
Add 500 ul Buffer PB
Apply to Qiaquick column and spin at 13,000 rpm for 1 min
Discard flow-thru
Add 750 ul Buffer PE, spin 60 sec
Aspirate flow thru and repeat
Aspirate flow thru and spin 1 min to dry column
Transfer to fresh tube
Add 30 ul Buffer EB to center of filter , let sit 1 min at RT
Spin 13,000 rpm for 1 min
Repeat elution step
Dry eluate in speed-vac

Pre-Hybridization:

Precipitate 50 ul salmon sperm DNA(1ug/ul) and resuspend in the same Hybridization mix as follows minus the probe. Apply to slide (create open well hybridization area with ~ 3 layers of rubber cement), create an airtight container using rubber gasket sealed with a glass slide and small clips.

incubate 37 C for 30 min.

Hybridization

Probe:
RESUSPEND PROBE IN 5 UL WATER

probe	5.0 ul	
Master Mix*	33.0 ul	
Poly A(10ug/ul)	2.0 ul	
tRNA (100 ug/ul)(Gibco)	1.0 ul	
10% SDS	1.0 ul	
COT 1 DNA (CONCENTRATED TO 10 UG/UL)		**8.0 UL**
(Gibco)	50 ul	

heat 70°, 10 min, ice 1 min
incubate 37°, 60 min

remove prehybridization mix and apply probe mix to slide, again seal with gasket, slide and small clips

incubate 37° on rocking table for 1-2 days

Washing:

wash 2X in 20 ml Formamide
4 ml 20XSSC
16 ml H20
40 ml at 45°, 15 min each

wash 2X in 1X PBS, 0.05% tween 20 at RT, 15 min each

Dapi counterstain→coverslip

*Master Mix

1g dextran sulfate
6 ml Formamide
1 ml 20 XSSC
7 ml

Assembly of microarrays for genome-wide measurement of DNA copy number

Published online: 30 October 2001, DOI: 10.1038/ng754

We have assembled arrays of approximately 2,400 BAC clones for measurement of DNA copy number across the human genome. The arrays provide precise measurement (s.d. of $\log_2$ ratios=0.05–0.10) in cell lines and clinical material, so that we can reliably detect and quantify high-level amplifications and single-copy alterations in diploid, polyploid and heterogeneous backgrounds.

Microarray-based comparative genomic hybridization (array CGH) provides a means to quantitatively measure DNA copy-number aberrations and to map them directly onto genomic sequence. Because arrays comprised of large-insert genomic clones such as BACs provide reliable copy-number measurements on individual clones[1,2], they are potentially useful for research and clinical applications in medical genetics and cancer. Preparation and spotting of BAC DNA is problematic, however, because (i) BACs are single-copy vectors (ii) the yield of DNA from BAC cultures is low compared to that from plasmid-bearing cultures and (iii) spotting high–molecular weight DNA at sufficient concentration to obtain a good ratio of signal to noise in the hybridizations may be difficult. To overcome these problems, we used ligation-mediated PCR[3] to generate representations of human and mouse BAC DNAs for spotting on arrays. We produced sufficient spotting solution (0.8 μg/μl DNA in 20% DMSO) from 1 ng of BAC DNA to make tens of thousands of arrays (see Web Note A for methods). The ratios we measured using arrays comprised of BAC representations are essentially identical to ratios previously reported for DNA from the same BACs[1]. Independently prepared DNA representations yield highly reproducible data (average variation of the linear ratios on individual clones from two independent preparations, 6.6%).

For copy-number assessment across the human genome, we printed 2,460 BAC and P1 clones in triplicate (approximately 7,500 elements) in a 12 mm×12 mm square (HumArray 1.14; see Web Table A and Web Fig. A). Each clone contains at least one STS, allowing linkage to the genome sequence. Cytogenetic mapping indicated that 2,298 of the arrayed clones are single copy[4,5]; these arrays thus provide average resolution of approximately 1.4 Mb across the genome. We have also assembled an array of approximately 1,300 clones for the mouse, which will be reported elsewhere. With the human arrays, we have obtained highly reproducible measurements over a wide dynamic range in cancer cell lines (see Web Tables B and C for analyses of COLO320, HCT116, HT29, MDA-MB-231, MDA-MB-453, MPE600, SW837 and T47D). These copy-number alterations ranged from homozygous deletions ($\log_2$ ratio<–2, HCT116 chromosome 16) to very high-level amplifications ($\log_2$ ratio>6, amplification of *CMYC*, COLO320). We also obtained nearly identical ratios (average s.d. of the $\log_2$ ratio=0.08) in three replicate hybridizations with BT474 cell line DNA, two labeled by random priming and one by nick translation, using an array of 1,777 clones (HumArray 1.11; see Web Table D and Web Fig. B).

To test our ability to measure single-copy changes (trisomies and monosomies), which is critical for applications in medical genetics and cancer, we measured 15 cell strains containing cytogenetically mapped partial or whole-chromosome aneuploidies (see Web Tables E–I). Figure 1 shows representative analyses, including detection of whole-chromosome gains (Fig. 1*a*), detection of a deletion (Fig. 1*b*) and its confirmation (Fig. 1*c*). The mean $\log_2$ ratios of trisomic chromosomal regions were

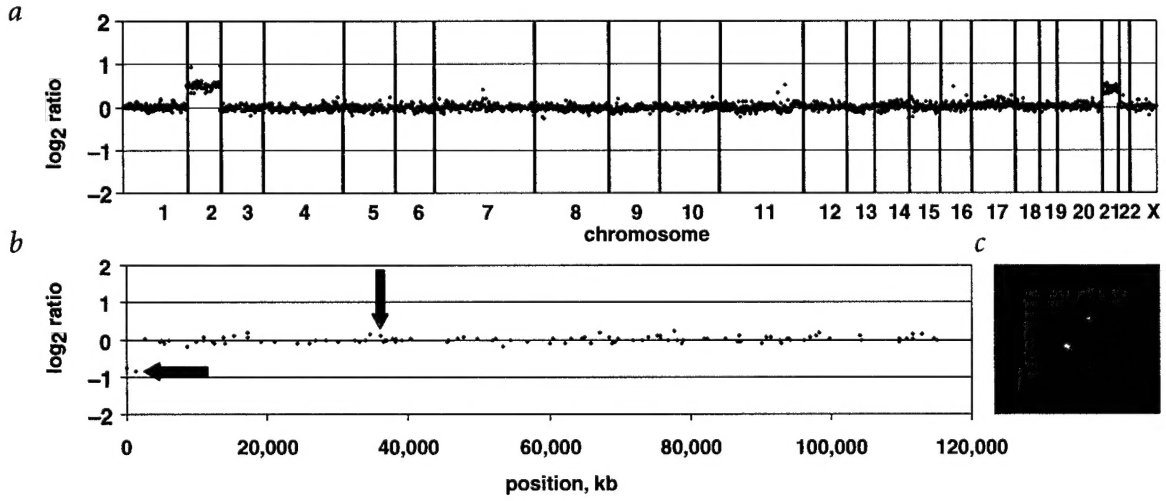

Fig. 1. Measurement of single-copy changes. *a*, Normalized copy-number ratios of a comparison of genomic DNA from cell strain GM03576 and from normal reference DNA (see Web Note for methods). Data are plotted as the mean $\log_2$ ratio of the triplicate spots for each clone normalized to the genome median $\log_2$ ratio. The BACs are ordered by position in the genome beginning at 1p and ending at Xq. Borders between chromosomes are indicated by vertical bars. Cytogenetic analysis indicates that this cell line is trisomic for chromosomes 2 and 21. *b*, Normalized copy-number variation of cell line GM03563 on BAC clones from chromosome 9. The mean $\log_2$ ratios of the triplicate spots normalized to the median $\log_2$ ratio for the genome are plotted relative to the position of the clones on the draft genomic sequence. The $\log_2$ ratio of approximately –1 indicates a single-copy deletion of the first two clones on chromosome 9. The standard deviation of the $\log_2$ ratios of the clones that are not deleted is 0.08. Colored arrows indicate clones hybridized to interphase nuclei in *c*. *c*, Confirmation of the deletion on 9p by fluorescent *in situ* hybridization to interphase GM03563 nuclei. The Texas red–labeled test clone RP11-28N06 included in the deletion (indicated by the red arrow in *b*) gave a single red hybridization signal and the FITC-labeled reference clone RP11-115L05 (indicated by the green arrow in *b*) gave two green signals.

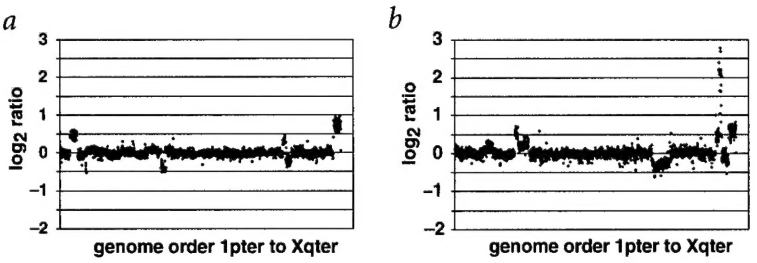

Fig. 2. Genome-wide copy-number variation in two breast tumors. *a* and *b*, Normalized fluorescence ratios for breast tumors. We labeled DNA extracted from sections of trimmed, frozen tumor specimens with Cy3–dCTP by random priming and hybridized it to the array together with normal male reference DNA (see Web Note for methods). We found low-level gains and losses in both tumors and a high-level amplification on chromosome 20q in *b*. The elevated X-chromosome ratios reflect the male–female difference in X-chromosome copy number. Ratios are plotted as in Fig. 1*a*.

0.49±0.05, compared to the ideal value of 0.58 for a 3/2 ratio. In female/male comparisons, the mean $\log_2$ ratios on the X chromosome were 0.72±0.08, compared to the expectation of 1.0. The underestimation of the magnitude of copy-number deviations most likely results from incomplete suppression of repetitive sequences or errors in background subtraction[1].

In principle, each clone may show a different relationship between ratio and copy number because of the differential ability to block its repetitive sequences. If so, we would expect that ratio differences among clones at the same copy number would become larger as the copy number departed farther from genome average. In the aneuploid cell lines, we found that the vast majority of the autosomal clones had the same response to copy-number changes, as the standard deviations of the $\log_2$ ratios for autosomal clones at 1, 2 or 3 copies were all 0.09. However, on the X chromosome, the standard deviation of the $\log_2$ ratios increased from 0.10 in male/male comparisons to 0.15 in female/female comparisons. Moreover, the ratio variations among X chromosome clones were very reproducible (see Web Fig. C), suggesting that the sequence characteristics of individual clones, possibly differing amounts of sequence shared with the Y chromosome, do have a measurable effect on X chromosome ratios.

We detected copy-number gains and losses (Fig. 2*a*,*b*; Web Table J) as well as amplifications (Fig. 2*b*) using DNA isolated from trimmed, frozen breast tumor tissue blocks. Many of the ratio changes are of smaller magnitude than would be expected for single-copy changes in diploid genomes. For example, the $\log_2$ ratios of 0.47±0.08 (Fig. 2*a*) and 0.32±0.07 (Fig. 2*b*) recorded for parts of the genome are less than the expected $\log_2$ ratio=0.58 for a copy-number ratio of 1.5. These ratios most likely reflect the presence of admixed normal cell DNA, tetraploid DNA content and/or tumor heterogeneity. In particular, the intermediate ratios indicating a gain of 16p and loss of 16q in Fig. 2*a* probably result from the presence of these aberrations in only a portion of the tumor cells. The magnitude of these easily discriminated ratio changes is well below the 'two-fold' level often considered to be the limit for significant differences in expression-array measurements, indicating the potential of array technology to provide very precise ratio measurements.

Previously, measuring DNA copy number using arrays assembled from representations of genomic clones prepared by other methods[6,7] resulted in highly variable ratios, so that detecting single-copy changes required averaging over several adjacent clones. In contrast, the arrays described here, produced from ligation-mediated PCR representations of the genomic clones, provide reliable data from individual clones, even in polyploid or heterogeneous specimens. This array CGH platform thus provides the performance required for potential clinical applications in medical genetic diagnosis and cancer.

Note: Supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Acknowledgments

We thank J. Flint, D. Ledbetter, C. Lese and Vysis, Inc. for telomere clones and clones containing certain named genes. This work was supported by NIH grants CA80314, CA83040, CA84118, HD17665 and CA58207, by California BCRP grant 2RB-0225 and by Vysis, Inc.

Antoine M. Snijders[1,2], Norma Nowak[4], Richard Segraves[1], Stephanie Blackwood[1,2], Nils Brown[1], Jeffrey Conroy[4], Greg Hamilton[1], Anna Katherine Hindle[1,2], Bing Huey[1], Karen Kimura[1], Sindy Law[1,2], Ken Myambo[1], Joel Palmer[1,2], Bauke Ylstra[1,2], Jingzhu Pearl Yue[1], Joe W. Gray[1,3], Ajay N. Jain[1–3], Daniel Pinkel[1,3] & Donna G. Albertson[1–3]

[1]Comprehensive Cancer Center, [2]Cancer Research Institute and [3]Department of Laboratory Medicine, University of California San Francisco, San Francisco, California 94143. [4]Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263. Correspondence should be addressed to D.G.A. (e-mail: albertson@cc.ucsf.edu).

Received 9 August; accepted 28 September 2001.

1. Pinkel, D. *et al. Nature Genet.* **20**, 207–211 (1998).
2. Albertson, D.G. *et al. Nature Genet.* **25**, 144–146 (2000).
3. Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R. & Riethmuller, G. *Proc. Natl Acad. Sci. USA* **96**, 4494–4499 (1999).
4. Cheung, V.G. *et al. Nature* **409**, 953–958 (2001).
5. Knight, S.J.L. *et al. Am. J. Hum. Genet.* **67**, 320–332 (2000).
6. Geschwind, D.H. *et al. Dev. Genet.* **23**, 215–29 (1998).
7. Lucito, R. *et al. Genome Res.* **10**, 1726–1736 (2000).